Un traitement hybride du vague textuel : du système expert VAGO à son clone neuronal

Benjamin Icard[(1)], Vincent Claveau[(2)], Ghislain Atemezing[(3)], Paul Égré[(1)]

(1) Institut Jean-Nicod, PSL University, 29 rue d'Ulm, 75005 Paris, France
(2) CNRS, IRISA, 263 Av. Général Leclerc, 35000 Rennes, France
(3) Mondeca, 18 rue de Londres, 75009 Paris, France

RÉSUMÉ

L'outil VAGO est un système expert de détection du vague lexical qui mesure aussi le degré de subjectivité du discours, ainsi que son niveau de détail. Dans cet article, nous construisons un clone neuronal de VAGO, fondé sur une architecture de type BERT, entraîné à partir des scores du VAGO symbolique sur un corpus de presse française (FreSaDa). L'analyse qualitative et quantitative montre la fidélité de la version neuronale. En exploitant des outils d'explicabilité (LIME), nous montrons ensuite l'intérêt de cette version neuronale d'une part pour l'enrichissement des lexiques de la version symbolique, et d'autre part pour la production de versions dans d'autres langues.

ABSTRACT

A hybrid treatment of textual vagueness: enriching the expert system VAGO with a neural clone

The VAGO tool is an expert system for lexical vagueness detection that also measures the degree of subjectivity of the speech, as well as its level of detail. In this paper, we build a neural clone of VAGO, based on a BERT-like architecture, trained on symbolic VAGO scores on a French press corpus (FreSaDa). The qualitative and quantitative analysis shows the fidelity of the neural version. By exploiting explainability tools (LIME), we then show the interest of this neural version for the enrichment of the lexicons of the symbolic version, and for the production of versions in other languages.

MOTS-CLÉS : Vague - Subjectivité - Précision - Détail - Hybridation - Explicabilité.

KEYWORDS: Vagueness - Subjectivity - Precision - Detail - Hybridization - Explainability.

1 Introduction

L'évaluation de la qualité d'un texte ou d'un discours s'avère nécessaire dans beaucoup d'applications, lesquelles reposent alors sur une définition propre de la notion de *qualité* (Štajner *et al.*, 2022). Pour un apprenant, on peut par exemple chercher à mesurer la complexité lexicale (Shardlow *et al.*, 2021) ou syntaxique des énoncés (Chen & Zechner, 2011). On peut également mesurer la complexité conceptuelle des termes et des énoncés à destination d'un jeune public (Štajner & Hulpuş, 2018) ou la compréhensibilité de formules administratives (François *et al.*, 2014). La lisibilité est une autre dimension largement explorée à partir d'indices divers (Collins-Thompson, 2014), allant jusqu'au rendu graphique pour certains troubles de la lecture (Rello & Baeza-Yates, 2016). S'agissant de la cohérence, on peut s'intéresser à l'enchaînement logique des énoncés, vu alors comme une tâche

d'*entailment* (Poliak, 2020). Notons enfin que ces questions de qualité de texte se posent également pour l'évaluation de systèmes de TAL produisant du texte (Celikyilmaz *et al.*, 2020), comme le résumé automatique, la traduction, la génération de texte, la simplification, etc.

Dans cet article, nous nous intéressons à une autre dimension de qualité du discours : la notion d'*informativité*. Là encore, des travaux existants explorent cette notion sous différents prismes. Tewari *et al.* (2020) utilisent par exemple des critères de cohésion syntaxique, alors que d'autres exploitent des critères de nouveauté par rapport à un ensemble de connaissances préexistantes (Chen *et al.*, 2018; Shibayama *et al.*, 2021). Pour notre part, nous nous inscrivons dans une veine de travaux où la mesure du degré d'informativité consiste à évaluer le caractère plus ou moins précis ou vague du discours (Van Deemter, 2010; Égré, 2018). Plus un discours est précis, plus il est apte à être infirmé s'il est faux (Popper, 1963) ; inversement plus un discours est vague, moins il se prête à la réfutation empirique et plus il est susceptible de véhiculer des informations subjectives (Égré & Icard, 2018). Pour automatiser l'évaluation de la qualité informationnelle d'un texte, il est donc utile de détecter des indices de vague comme de précision.

Dans ce but, nous proposons ici de combiner deux approches. D'une part, nous faisons appel à un système expert de détection et de mesure du vague lexical, l'outil VAGO (Guélorget *et al.*, 2021; Icard *et al.*, 2022). De l'autre, nous proposons d'en créer une version neuronale afin de tester comme d'enrichir les performances du système expert. L'un des enjeux de cette méthode est d'étendre à d'autres langues que le français et l'anglais les résultats du système expert. Un autre est d'avancer dans la maîtrise de méthodes hybrides de traitement du langage.

2 L'outil symbolique VAGO

2.1 Typologie du vague et mesure du niveau de détail

VAGO mesure le vague et la subjectivité des documents à partir d'une base de données lexicales en français et en anglais (Atemezing *et al.*, 2021). Fondée sur une typologie issue de (Égré & Icard, 2018), cette base de données propose un inventaire des termes vagues en quatre catégories : vague d'approximation (V_A), vague de généralité (V_G), vague de degré (V_D) et vague combinatoire (V_C).

Le vague d'approximation concerne principalement des modificateurs comme *"environ"*, qui rendent moins strictes les conditions de vérité de l'expression qu'ils modifient. Le vague de généralité comprend des déterminants comme *"certains"*, ainsi que des modificateurs comme *"au plus"*. Contrairement aux expressions d'approximation, ces dernières ont des conditions de vérité précises. La classe des expressions relevant du vague de degré et du vague combinatoire (Alston 1964) comprend principalement des adjectifs unidimensionnels d'une part (tels que *"grand"*, *"vieux"*) et des adjectifs multidimensionnels d'autre part (comme *"beau"*, *"intelligent"*, *"bon"*, *"qualifié"*). Les expressions de type V_A et V_G sont en outre traitées comme des expressions *factuelles*, et les expressions de type V_D et V_C comme des expressions *subjectives* (Kennedy, 2013; Verheyen *et al.*, 2018; Solt, 2018).

Selon les règles de calcul de ratio détaillées en 2.2, dans la version princeps de VAGO il suffit qu'une phrase contienne au moins un marqueur de vague, respectivement de vague subjectif, pour que VAGO la considère comme vague, respectivement comme subjective[1]. Cependant, cette version n'offre pas

[1]En plus de ces différents ratios, VAGO s'appuie sur plusieurs règles de modulation des scores de vague en fonction du

de mesure *positive* pour la précision du discours: une phrase est jugée précise si elle ne contient aucun marqueur vague ; un texte est jugé précis s'il ne contient aucune phrase vague. À titre d'exemple, la version en ligne de VAGO princeps attribuera des scores de vague et de subjectivité identiques, égaux à 1, aux deux phrases suivantes[2]:

(a) "Roi de Naples de 1806 à 1808, puis d'Espagne de 1808-1813, il est un personnage **important** du dispositif que met en place Napoléon pour asseoir la souveraineté de la France sur l'Europe continentale".

(b) "Pour guérir **rapidement** de la Covid-19 il faut prendre une **excellente** décoction de plantes".

Les phrases (a) et (b) contiennent chacune au moins un marqueur de vague également vecteur de subjectivité: un seul pour (a) avec "*important*" et deux pour (b) avec "*rapidement*" et "*excellente*". Ces phrases sont donc identiquement vagues/subjectives selon VAGO princeps. Intuitivement, pourtant, la phrase (a) qui contient neuf entités nommées (termes soulignés dans la phrase) est plus informative que la phrase (b) qui ne contient qu'une seule entité nommée ("*Covid-19*") et renferme donc moins de détails que (a). Pour combler cette lacune, la version princeps de VAGO est enrichie ici d'un score de détail fondée sur la part relative des entités nommées comparées aux expressions vague.

2.2 Scores VAGO : vague, subjectivité, détail

Actuellement, VAGO permet de mesurer le score de vague, de subjectivité et le niveau de détail de documents anglais ou français. La détection du vague et de la subjectivité s'appuie sur la base de donnée princeps comportant actuellement 1 640 termes dans les deux langues (Atemezing *et al.*, 2022), répartis comme suit par catégorie de vague : $|V_A| = 9$, $|V_G| = 18$, $|V_D| = 43$ et $|V_C| = 1570$. Dans le cas du niveau de détail, la détection se fonde sur l'identification des entités nommées par spaCy[3] (personnes, localités, indications temporelles, institutions, nombres). Pour chaque mesure, les marqueurs caractéristiques sont détectés et notés des mots vers les phrases, puis des phrases vers les textes.

Pour une phrase ϕ donnée, son *score de vague* est défini comme le rapport entre le nombre de mots vagues dans ϕ et le nombre total de mots dans la phrase N_ϕ :

$$R_{vague}(\phi) = \frac{\overbrace{|V_D|_\phi + |V_C|_\phi}^{subjectif} + \overbrace{|V_A|_\phi + |V_G|_\phi}^{factuel}}{N_\phi} \quad (1)$$

où $|V_A|_\phi$, $|V_G|_\phi$, $|V_D|_\phi$ et $|V_C|_\phi$ représentent le nombre de termes dans ϕ relevant de chacune des quatre catégories de vague (approximation, généralité, vague de degré et vague combinatoire). Plus finement, le *score de subjectivité* d'une phrase est calculé comme le rapport entre les expressions vagues subjectives et le nombre total de mots de la phrase. On peut calculer un score de vague factuel

contexte (voir Icard *et al.*, 2022).

[2]La phase (a) est tirée de l'article Wikipédia sur Joseph Bonaparte, tandis que la phrase (b) est inspirée d'une fausse nouvelle ou "*fake news*".

[3]https://spacy.io/

identiquement avec les expressions de généralité et d'approximation (cf. sections 3 et 4) :

$$R_{subjectif}(\phi) = \frac{|V_D|_\phi + |V_C|_\phi}{N_\phi} \qquad R_{vaguefactuel}(\phi) = \frac{|V_A|_\phi + |V_G|_\phi}{N_\phi} \tag{2}$$

Le *score de détail* d'une phrase peut être défini comme le ratio $R_{dtail}(\phi) = \frac{|P|_\phi}{N_\phi}$, où $|P|_\phi$ désigne le nombre d'entités nommées de la phrase (termes référentiels). Par extension, si $|V|_\phi$ désigne le nombre de termes vagues d'une phrase (toutes catégories confondues), on définit le *score de détail/vague* d'une phrase comme la part relative des entités nommées, soit :

$$R_{détail/vague}(\phi) = \frac{|P|_\phi}{|P|_\phi + |V|_\phi} \tag{3}$$

Dans l'exemple précédent, on peut vérifier que $R_{détail/vague}(\text{a}) = 9/10$, alors que $R_{détail/vague}(\text{b}) = 1/3$, ce qui donne une meilleure mesure du caractère plus informatif de (a).

Pour des ensembles de phrases, ou textes T, les scores de vague de T (respectivement de vague subjectif, ou de vague factuel) sont définis comme la proportion de phrases de T dont le score de vague (resp. subjectif, ou factuel) sont non nuls. Le score de détail/vague de T, lui, est défini comme la moyenne des ratios $R_{détail/vague}$ de chacune des phrases de T.

2.3 Implémentation des mesures

En terme d'ingénierie, VAGO s'appuie sur GATE (Cunningham, 2002) pour le traitement des corpus. L'algorithme exploite également l'annotateur de contenu sémantique CA-Manager (Cherfi *et al.*, 2013) qui sert à extraire des connaissances à partir de données non structurées. En l'état, VAGO détecte automatiquement la langue du corpus (anglais ou français) à l'aide du module TextCat[4]. La version neuronale de VAGO présentée en section 3 utilise spaCy pour la détection des entités nommées.

L'outil VAGO en ligne, disponible sur le site de Mondeca[5], présente les fonctionnalités de VAGO dans sa version princeps. Le site propose une interface graphique pour mesurer les scores de vague et de subjectivité de textes sous la forme de deux baromètres. Le premier baromètre représente le degré de vague d'un texte (défini comme $R_{vague}(T)$) tandis que le second baromètre indique le degré auquel le texte rapporte une opinion plutôt qu'un fait, autrement dit la proportion de vocabulaire subjectif au sein du texte (défini comme $R_{subjectif}(T)$).

2.4 Test de VAGO sur la presse française

VAGO a été testé sur le corpus français "FreSaDa"[6] (Ionescu & Chifu, 2021) composé de 11 570 articles de presse répartis en deux classes supposées homogènes : 5 648 articles "réguliers" et provenant de la presse française généraliste, sans faire de présupposition sur leur vérité ou leur

[4] https://www.let.rug.nl/vannoord/TextCat/index.html
[5] https://research.mondeca.com/demo/vago/
[6] https://github.com/adrianchifu/FreSaDa

fausseté, versus 5 922 articles “satiriques” et explicitement faux à ce titre. Au sein du corpus total, VAGO a traité 10 969 articles des 11 570 articles initiaux, — les 601 articles restants ayant été écartés car ils ne relèvent pas d’un format adéquat pour être traité par l’outil (mots isolés, mots-clés, phrases incomplètes, etc.). Les résultats fournis par VAGO sont rapportés en Figure 1.

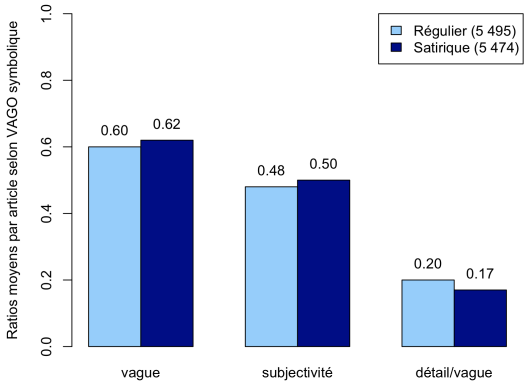


Figure 1: Ratios moyens par article du corpus FreSaDa selon VAGO.

Selon VAGO, les articles du corpus satirique sont significativement *plus vagues* ($p = 4.99 \times 10^{-11}$), *plus subjectifs* ($p = 1.69 \times 10^{-9}$) et *moins détaillés* ($p = 3.36 \times 10^{-22}$) que les articles du corpus de presse régulier (scores calculés par textes ; t-tests bilatéraux, $\alpha = 0.05$, avec correction de Bonferroni). Ces résultats sont conformes aux attentes et corroborent les résultats obtenus antérieurement avec VAGO sur des textes en anglais (Guélorget *et al.*, 2021). En outre, des expériences non rapportées dans cet article montrent que les ratios calculés par VAGO, utilisés en entrée d’un classifieur, permettent de distinguer les documents selon leur catégorie *légitime* ou *biaisée* avec une grande précision.

3 La version neuronale VAGO-N

À partir d’une architecture BERT (Devlin *et al.*, 2018), nous présentons ici une version neuronale VAGO-N de la version symbolique VAGO décrite précédemment. Cette version neuronale vise à dépasser certaines limites de VAGO et à ouvrir la voie à plusieurs développements que nous présentons dans la section suivante.

3.1 Apprentissage du clone VAGO-N

L’architecture BERT est associée à une couche de régression ainsi qu’à une fonction de perte MSE afin de prédire un score associé à un texte ; nous testons les deux type de vague : subjectif ou factuel. Par souci de complétude, nous testons également la prédiction du score $R_{détail}$, mais ce score peut être plus simplement calculé à partir d’un système de reconnaissance d’entités nommées ; nous n’y revenons donc pas dans les expériences suivantes. Comme pour une tâche de distillation, VAGO est donc utilisé pour associer un score de vague aux phrases d’un corpus et entraîner ainsi un système neuronal.

Dans les expériences rapportées, 106 000 phrases sont tirées aléatoirement des 10 969 articles du corpus FreSaDa traités par VAGO, et sont classiquement divisées en jeu d'entraînement (85 000 phrases) et de test (21 000 phrases). Nous utilisons un modèle RoBERTa Large (*Batch Size*=30 ; *Learning Rate*=1e-6 ; *Epochs*=20) ; des expériences non détaillées ici avec un modèle CamemBERT (Martin *et al.*, 2019) fournissent des résultats légèrement inférieurs.

Les performances sont rapportées en Table 1 avec les mesures standard de régression : l'erreur quadratique moyenne (RMSE), le coefficient de détermination (R^2), l'erreur absolue moyenne (MAE) et l'erreur absolue médiane (MedAE). Toutes ces mesures montrent que VAGO−N réplique avec une grande précision les scores du VAGO symbolique. La tâche de détection de la subjectivité semble un peu plus difficile que celle du vague factuel.

	RMSE	R^2	MAE	MedAE
vague subjectif	0.022063	0.859897	0.014518	0.009488
vague factuel	0.008745	0.949339	0.004124	0.001730
détail/vague	0.097008	0.882543	0.051396	0.012367

Table 1: Résultats de régression de VAGO−N sur les phrases du corpus FreSaDa (français) pour les scores de vague subjectif, de vague factuel et de détail par rapport au vague.

3.2 Comparaison des versions de VAGO

L'évaluation quantitative précédente indique que VAGO−N réplique assez fidèlement le comportement général de VAGO. Il est intéressant de vérifier de manière plus qualitative que cette version neuronale s'appuie bien sur les mêmes indices lexicaux que la version symbolique.

Pour cela, nous utilisons l'outil d'explicabilité LIME (Ribeiro *et al.*, 2016). Appliqué aux sorties de VAGO−N, LIME permet d'identifier les tokens qui contribuent le plus (ou le moins) au score de vague d'un texte donné. Dans le cas d'une phrase en français, un exemple de sortie de LIME concernant le cas du vague subjectif est fourni en Figure 2. Avec cet outil, nous examinons les cas où les prédictions (scores de vague) de VAGO−N divergent le plus de celles de VAGO.

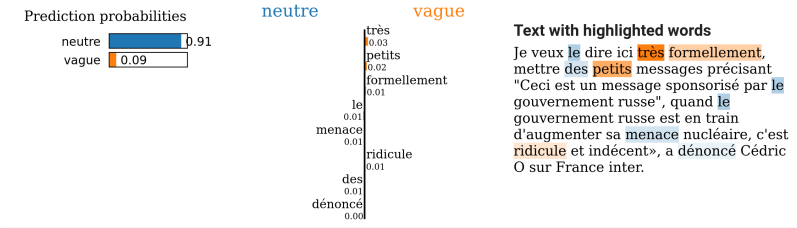


Figure 2: Exemple de sortie de LIME sur une phrase du corpus FreSaDa traitée par VAGO−N. La catégorie notée neutre correspond à la catégorie inverse du vague, contribuant négativement au score de vague.

L'étude de ces cas d'erreurs fait ressortir plusieurs points. Concernant les écarts de prédiction entre VAGO et VAGO−N sur le vague factuel, la très grande majorité des termes identifiés comme contribuant

le plus fortement à la prédiction de VAGO−N figurent déjà au sein du lexique français de VAGO, aussi bien pour le vague de généralité V_G (e.g. *"tout/tous/toutes", "jamais", "ou", "général", "quelques", "certains/certaines"*), que pour le vague d'approximation (*"environ", "presque"*). Les indices de vague factuel repérés sont corrects mais leur poids dans le score final de VAGO−N diffère du calcul effectué par le VAGO initial qui n'établit pas de pondération. Cette différence de poids attribuée par VAGO−N peut résulter de mots d'autres catégories morpho-syntaxiques (les lexiques de VAGO se focalisant sur les adjectifs et adverbes) qui viennent amplifier ou amoindrir le score de vague factuel résultant.

De manière similaire, dans le cas du vague subjectif, LIME appliqué à VAGO−N relève des adjectifs et des termes d'outrance déjà présents dans le lexique existant, soit au sein du vague combinatoire pour majorité (e.g. *"négatif", "affirmatif", "intéressant", "fortement", "difficile", "probablement", "vrai", "stupide", "vraiment"*), soit au sein du vague de degré (*"petit"*). LIME identifie également d'autres adjectifs porteurs de vague combinatoire qui ne figurent pas encore dans le lexique mais sont voués à y figurer (e.g. *"durable", "particulièrement", "ringard", "actuel"*), avec toutefois des exceptions ("*sabbatique*") ; nous y revenons dans la section suivante.

4 Extensions des approches VAGO

La section précédente montre qu'il est possible de construire un équivalent de la version française de VAGO par apprentissage. Cela permet d'explorer plusieurs développements que nous présentons dans les deux sous-sections suivantes : l'enrichissement de la base lexicale au coeur de VAGO, et la construction de systèmes de détection du vague pour d'autres langues.

4.1 Validation et enrichissement du VAGO symbolique

Comme nous l'avons vu précédemment, pour chaque token t dans un texte, LIME fournit un score de contribution de t à la prédiction de vague (subjectif ou factuel) du texte que nous notons $c_{occ}(t)$. Dans le cas d'une phrase, plus un terme t reçoit un score $c_{occ}(t)$ élevé, plus il contribue positivement au score de vague de cette phrase.

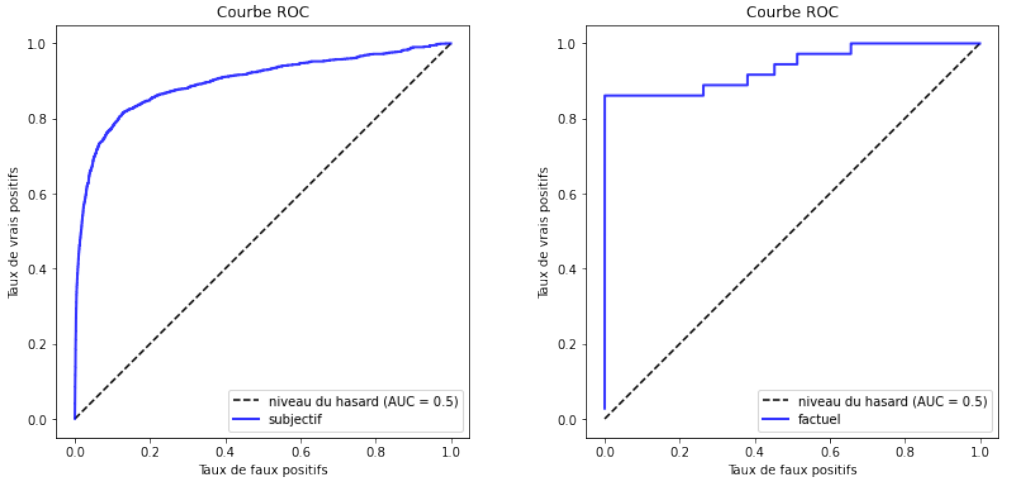
En appliquant LIME aux phrases des 10 969 articles du corpus FreSaDa traités par VAGO et exploités dans VAGO−N, nous collectons les scores de contribution c_{occ} de toutes les occurrences de tous les tokens figurant au sein de ces phrases. Pour obtenir un score global $c_{tok}(t)$ par token t, nous sommons et normalisons les c_{occ} par le nombre total d'occurrences de chaque token noté ici $|occ_t|$: $c_{tok}(t) = \frac{1}{|occ_t|} \sum_{o \in occ_t} c_{occ}(o)$. Notre hypothèse est que des termes du lexique de VAGO devraient se retrouver prioritairement parmi les tokens recevant les c_{tok} les plus élevés. À cet égard, nous calculons les précisions statistiques P@i (vrais positifs/(vrais+faux positifs)) sur la liste des tokens ordonnées par c_{tok} décroissants. À noter qu'un token est pris en compte s'il est une flexion d'un terme du lexique VAGO. Les résultats sont recensés en Table 2.

La Figure 3 présente la courbe ROC reliant le score c_{tok} à la présence au sein du lexique VAGO. Ces résultats établissent le bien-fondé de notre hypothèse. Ainsi, VAGO−N, bien qu'entraîné uniquement sur des phrases et leur score, est capable de reconstruire le lexique au cœur de la version symbolique.

De plus, nous avons examiné les 100 tokens détenteurs des c_{tok} les plus élevés pour le vague

	P@5	P@10	P@20	P@30	P@100	P@200
vague subjectif	1.00	1.00	0.95	0.93	0.81	0.79
vague factuel	1.00	1.00	1.00	0.93	0.31	0.16

Table 2: Comparaison de la précision obtenue à différents seuils de la liste de tokens du lexique VAGO ordonnée par c_{tok}.



(a) vague subjectif

(b) vague factuel

Figure 3: Courbe ROC de c_{tok} comme indicateur de présence dans le lexique français de VAGO ; a) correspond au vague subjectif, b) au vague factuel.

subjectif : outre les 81 bien présents dans le lexique, quelques formes verbales sont listées. Bien que considérés comme des faux positifs (les lexiques VAGO actuels ne recensant que les adjectifs et adverbes), leur pertinence peut être discutée. Dans les tokens restants, sept mots absents ont été validés comme pertinents et méritant d'être intégrés dans les lexiques. Cette liste contient quatre adverbes (*"également", "seulement", "particulièrement", "clairement"*) pour lesquels la base de données VAGO contient les adjectifs racines dans deux cas (*"particulier", "clair"*) ; un verbe d'action (*"faire"*) ; un adjectif pouvant également être un nom (*"droit/droite"*) ; et un nom (*"nombre"*). On note également la détection de formes non standard de termes présents dans le lexique (*"difficille", "pauv"*), illustrant la robustesse de l'approche neuronale sur du texte bruité (coquille, abréviation...). Les résultats obtenus valident le clone neuronal VAGO-N qui retrouve les indices lexicaux du système expert VAGO tout en permettant d'en identifier de nouveaux ou des formes non standard.

4.2 Développement de VAGO-N multilingues

Le développement de versions symboliques de VAGO pour d'autres langues implique de disposer de lexiques de vague dans les langues cibles. Cependant, la traduction automatique de ces lexiques n'est pas possible en raison du caractère idiomatique et hors contexte des listes d'expressions en jeu.

	RMSE	R^2	MAE	MedAE
vague subjectif	0.031801	0.708915	0.022865	0.016807
vague factuel	0.016990	0.808772	0.009582	0.004172

Table 3: Résultats de régression de VAGO−N sur les phrases du corpus FreSaDa (traduites automatiquement en anglais) concernant les scores de vague subjectif et de vague factuel.

	P@5	P@10	P@20	P@30	P@100	P@200
vague subjectif	0.80	0.90	0.90	0.93	0.90	0.84
vague factuel	1.00	0.80	0.55	0.50	0.26	0.14

Table 4: Comparaison de la précision obtenue à différents seuils de la liste de tokens ordonnée par c_{tok} en fonction du lexique VAGO.

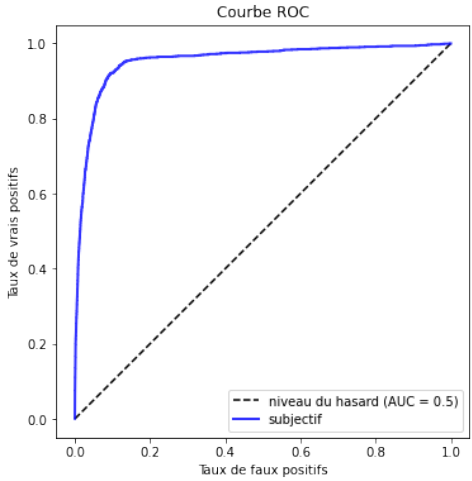
Cela étant, il est possible de traduire le jeu d'entraînement de VAGO−N en faisant l'hypothèse suivante : les scores de vague, en particulier de vague subjectif et de vague factuel, sont conservés de la langue source vers la langue cible. Pour commencer, le corpus FreSaDa a été traduit du français vers l'anglais en utilisant le modèle Helsinki-NLP/opus-mt-fr-en[7] (Tiedemann & Thottingal, 2020). Ensuite, VAGO−N a été entraîné à prédire les scores de vague subjectif et de vague factuel sur ce corpus en anglais (en utilisant les mêmes hyper-paramètres que pour l'entraînement de VAGO−N sur le français). Les résultats de régression sont similaires à ceux obtenus pour le français, et présentés en Table 3.

En appliquant la même approche qu'en sous-section 4.1, nous isolons la liste des tokens ordonnées par c_{tok} décroissants, puis la comparons au lexique anglais de VAGO servant ainsi de vérité-terrain. La précision de cette liste mesurée à différents seuils est rapportée en Table 4. Les courbes ROC correspondantes sont présentées en Figure 4.

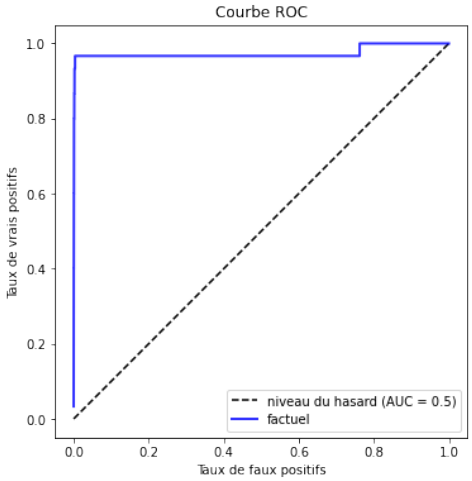
Nous collectons également les 100 termes anglais les plus porteurs de vague selon VAGO−N. Ces termes sont comparés à ceux du lexique anglais de la version symbolique : 90 termes figurent déjà au sein du lexique anglais de VAGO.

Parmi les termes de rang le plus élevé parmi les 100 premiers qui ne figurent pas dans VAGO, on trouve cinq adjectifs ou adverbes ayant vocation à figurer dans le vague combinatoire (*"likely", "full", "complicated", "frankly", "enough"*), un modal (*"must"*), qu'il est également sensé d'intégrer étant donné la présence de *"should"* dans le lexique VAGO. Quatre termes en revanche ne relèvent pas clairement du vague (*"course", "lost", "lose" et "finally"*), sauf éventuellement le premier (occurrences de "of course" dont l'usage est subjectif). Parmi les 100 termes suivants, tous les termes qui ne figurent pas dans VAGO sont des adjectifs pouvant figurer dans la catégorie V_C (*"worse", "complex"*, etc.).

[7] https://huggingface.co/Helsinki-NLP/opus-mt-fr-en



(a) vague subjectif



(b) vague factuel

Figure 4: Courbe ROC de c_{tok} comme indicateur de présence dans le lexique anglais de VAGO ; a) correspond au vague subjectif, b) au vague factuel à partir de textes traduits en anglais.

5 Conclusion

Dans cet article, nous sommes partis d'un système expert de détection du vague, l'outil VAGO, qui associe à des textes des scores de vague, de subjectivité, et de niveau de détail/vague, et nous avons ensuite créé un clone neuronal de VAGO fondé sur BERT. À la différence de VAGO, VAGO−N est entraîné uniquement à partir des scores de VAGO, sans connaître le lexique sous-jacent à cette version symbolique. En utilisant LIME, il apparaît que les termes dont la contribution aux scores du VAGO−N sont les plus élevés sont ou bien des termes figurant dans VAGO, ou des termes ayant majoritairement vocation à y figurer. Ce résultat suggère que les décisions de VAGO−N s'expliquent dans une large mesure à partir des items lexicaux identifiés par le VAGO symbolique. On a pu monter l'intérêt de ces croisements entre système symbolique et neuronal : une fois appris, VAGO−N permet de compléter les lexiques du VAGO symbolique, il permet de produire facilement des versions neuronales dans d'autres langues, et rend possible des versions symboliques dans ces langues en produisant les lexiques nécessaires.

Beaucoup de pistes restent à explorer. Il serait notamment intéressant de comparer les indices lexicaux trouvés par LIME dans un classifieur entrainé à distinguer directement les types de documents de FreSada (ou d'autres corpus pour la détection de Fake News) et ceux obtenus par notre approche, que nous espérons plus génériques. Une autre piste que nous souhaitons explorer pour mesurer la part de généricité de notre approche serait de masquer les entités nommées dans un texte pour voir si les scores de VAGO−N restent stables avant et après ce masquage, et pour déterminer la part revenant proprement au lexique VAGO (et qui ne contient pas d'entités nommées) dans la décision du VAGO−N. Le système expert VAGO est par définition un système rigide qui ne différencie par le caractère plus ou moins vague d'un terme selon le contexte. Prenons un terme comme *"affirmatif"* : le fait pour une phrase d'être *"affirmative"* n'est pas vague, en revanche si une personne est dite *"très affirmative"*,

alors *"affirmatif"* revêt un sens vague et évaluatif. À la différence de VAGO, VAGO−N est capable de différencier les scores de contribution de ces différentes occurrences d'un terme. Parmi les questions que nous réservons à un examen ultérieur, il serait particulièrement fructueux d'examiner la variabilité des scores de termes vagues selon les différents contextes dans lesquels ils apparaissent. La question se pose plus généralement pour BERT, de savoir si les plongements des termes vagues manifestent une dispersion plus grande que les plongements des termes fonctionnels et précis du lexique.

Remerciements

Nous remercions deux rapporteurs pour leur commentaires, ainsi que Guillaume Gravier (CNRS – IRISA) pour son retour et ses suggestions sur la version préliminaire de cet article. Ce travail a été réalisé dans le cadre du programme HYBRINFOX (ANR-21-ASIA-0003) (CNRS, IRISA, Mondeca, Airbus). PE et BI remercient également le programme ANR-17-EURE-0017 (FrontCog), et PE le programme PLEXUS (Marie Skłodowska-Curie Action, Horizon Europe Research and Innovation Programme, grant agreement n°101086295).

6 Bibliographie

ALSTON W. P. (1964). *Philosophy of Language*. Prentice Hall.

ATEMEZING G., ICARD B. & ÉGRÉ P. (2021). Multilingual gazetteers to detect vagueness in textual documents. DOI : 10.5281/zenodo.4718530.

ATEMEZING G., ICARD B. & ÉGRÉ P. (2022). Vague Terms in SKOS to detect vagueness in textual documents. DOI : 10.5281/zenodo.4718530.

CELIKYILMAZ A., CLARK E. & GAO J. (2020). Evaluation of text generation: A survey. DOI : 10.48550/ARXIV.2006.14799.

CHEN D., MA S., YANG P. & SUN X. (2018). Identifying high-quality chinese news comments based on multi-target text matching model. DOI : 10.48550/ARXIV.1808.07191.

CHEN M. & ZECHNER K. (2011). Computing and evaluating syntactic complexity features for automated scoring of spontaneous non-native speech. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, p. 722–731, Portland, Oregon, USA: Association for Computational Linguistics.

CHERFI H., COSTE M. & AMARDEILH F. (2013). CA-manager: a middleware for mutual enrichment between information extraction systems and knowledge repositories. In *4th workshop SOS-DLWD*, p. 15–28.

COLLINS-THOMPSON K. (2014). Computational assessment of text readability: A survey of current and future research. *ITL - International Journal of Applied Linguistics*, **165**, 97–135. DOI : 10.1075/itl.165.2.01col.

CUNNINGHAM H. (2002). GATE, a general architecture for text engineering. *Computers and the Humanities*, **36**(2), 223–254. DOI : 10.1023/A:1014348124664.

DEVLIN J., CHANG M.-W., LEE K. & TOUTANOVA K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv:1810.04805 [cs]*.

ÉGRÉ P. (2018). *Qu'est-ce que le vague?* Vrin.

ÉGRÉ P. & ICARD B. (2018). Lying and vagueness. In J. MEIBAUER, Éd., *Oxford Handbook of Lying*. OUP.

FRANÇOIS T., BROUWERS L., NAETS H. & FAIRON C. (2014). AMESURE: a readability formula for administrative texts (AMESURE: une plateforme de lisibilité pour les textes administratifs) [in French]. In *Proceedings of TALN 2014 (Volume 2: Short Papers)*, p. 467–472, Marseille, France: Association pour le Traitement Automatique des Langues.

GUÉLORGET P., ICARD B., GADEK G., GAHBICHE S., GATEPAILLE S., ATEMEZING G. & ÉGRÉ P. (2021). Combining vagueness detection with deep learning to identify fake news. In *Proceedings of 24th International Conference on Information Fusion*, p.8.

ICARD B., ATEMEZING G. & ÉGRÉ P. (2022). VAGO: un outil en ligne de mesure du vague et de la subjectivité. In *Conférence Nationale sur les Applications Pratiques de l'Intelligence Artificielle (PFIA 2022)*, p. 68–71.

IONESCU R.-T. & CHIFU A.-G. (2021). FreSaDa: A french satire data set for cross-domain satire detection. In *The International Joint Conference on Neural Network, IJCNN 2021*, IJCNN2021.

KENNEDY C. (2013). Two sources of subjectivity: Qualitative assessment and dimensional uncertainty. *Inquiry*, **56**(2-3), 258–277.

MARTIN L., MULLER B., SUÁREZ P. J. O., DUPONT Y., ROMARY L., DE LA CLERGERIE É. V., SEDDAH D. & SAGOT B. (2019). CamemBERT: a tasty french language model. *arXiv preprint arXiv:1911.03894*.

POLIAK A. (2020). A survey on recognizing textual entailment as an NLP evaluation. In *Proceedings of the First Workshop on Evaluation and Comparison of NLP Systems*, p. 92–109, Online: Association for Computational Linguistics. DOI : 10.18653/v1/2020.eval4nlp-1.10.

POPPER K. R. (1963). *Conjectures and refutations: the growth of scientific knowledge*. New York: Basic Books.

RELLO L. & BAEZA-YATES R. (2016). The effect of font type on screen readability by people with dyslexia. *ACM Trans. Access. Comput.*, **8**(4). DOI : 10.1145/2897736.

RIBEIRO M. T., SINGH S. & GUESTRIN C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, p. 1135–1144, New York, NY, USA: Association for Computing Machinery. DOI : 10.1145/2939672.2939778.

SHARDLOW M., EVANS R., PAETZOLD G. H. & ZAMPIERI M. (2021). SemEval-2021 task 1: Lexical complexity prediction. In *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021)*, p. 1–16, Online: Association for Computational Linguistics. DOI : 10.18653/v1/2021.semeval-1.1.

SHIBAYAMA S., YIN D. & MATSUMOTO K. (2021). Measuring novelty in science with word embedding. *PLoS ONE*, **16**(7). DOI : 10.1371/journal.pone.0254034.

SOLT S. (2018). Multidimensionality, subjectivity and scales: Experimental evidence. In E. CASTROVIEJO, L. MCNALLY & G. SASSOON, Éds., *The Semantics of Gradability, Vagueness, and Scale Structure*, p. 59–91. Springer.

ŠTAJNER S. & HULPUŞ I. (2018). Automatic assessment of conceptual text complexity using knowledge graphs. In *Proceedings of the 27th International Conference on Computational Linguistics*, p. 318–330, Santa Fe, New Mexico, USA: Association for Computational Linguistics.

ŠTAJNER S., SAGGION H., FERRÉS D., SHARDLOW M., SHEANG K. C., NORTH K., ZAMPIERI M. & XU W., Éds. (2022). *Proceedings of the Workshop on Text Simplification, Accessibility, and Readability (TSAR-2022)*, Abu Dhabi, United Arab Emirates (Virtual). Association for Computational Linguistics.

TEWARI M., BENSCH S., HELLSTRÖM T. & RICHTER K.-F. (2020). Modelling grice's maxim of quantity as informativeness for short text. In ∴, p. 1–7.

TIEDEMANN J. & THOTTINGAL S. (2020). OPUS-MT — Building open translation services for the World. In *Proceedings of the 22nd Annual Conferenec of the European Association for Machine Translation (EAMT)*, Lisbon, Portugal.

VAN DEEMTER K. (2010). *Not exactly: In praise of vagueness*. OUP Oxford.

VERHEYEN S., DEWIL S. & ÉGRÉ P. (2018). Subjectivity in gradable adjectives: The case of *tall* and *heavy*. *Mind & Language*, **33**(5), 460–479.